Achromatic Color Constancy and Lightness Functions
For a Complex Pattern of Luminance

Kaoru NOGUCHI and Masanori MOTOKI

Chiba University Japan Automobile Research Institute

Illumination of a cross-shaped pattern composed of five black to white squares on a white background was varied over 2.6 log units. The task given to 10 Ss was to make achromatic color namings and to make numerical judgments of lightness for each of the five squares in the cross as well as the background. Achromatic color constancy was almost perfect for the region of the highest contrast ratio of surround to focal luminance and less perfect for the regions with lower contrast ratios. Lightness judgments were found to change as a curvilinear function (concave downward) for a log transformation of illumination. Linearly decreasing functions were exceptional. It can be concluded, therefore, that the theoretical claim for the generality of negative functions for lightness-illumination relationships should be reconsidered.

Key words: achromatic color constancy, brightness-and-lightness constancy, contrast ratio, lightness-illumination relationships.

One of controversial issues in the research area of brightness contrast and constancy concerns the phenomenon that is being judged and measured. Some investigators explicitly set brightness as the phenomenon. Others explicitly set lightness as the phenomenon. Still others simply instruct their subjects to make an equality match (in brightness and/or lightness). Most of vision researchers seem to set brightness as the phenomenon and they tend to make no distinction between brightness and lightness. Phenomenology-oriented perceptionists set lightness as the phenomenon, making clearer distinctions between the two perceptual dimensions.

Classical studies have been mostly carried out in a simple situation such as an arrangement composed of a test-field, an inducing-field, and a dark surround. It is not clear, however, that such a simple stimulus arrangement is appropriate for studying the problem of perceptual constancy where the distinction between brightness and lightness cannot be ignored. As Flock and Noguchi (1970) proposed, brightness constancy really refers to the fact that the achromatic colors of surfaces supposedly remain phenomenally invariant over changes of illumination in a complex environment. It is characteristic in a well-organized visual environment that there are clearly different modes of appearance of colors. As far as the problem of constancy is concerned, therefore, more research should be required to make clear what phenomenon is being judged and measured using a complex pattern.

There have been several attempts to investigate brightness constancy in a complex field. The first attempt was made by Jameson and Hurvich (1961). In their experiment a cross-shaped configuration composed of five different
grays was presented against a medium gray background under three levels of illumination covering a range of 1.1 log units. The results showed that as illuminance was increased, the high reflectance regions yielded positive functions; the low reflectance region yielded a negative function; and some region in between yielded a zero function (perfect constancy). In order to find a similar form of brightness functions, with particular reference to the occurrence of negative slopes, Flock and Noguchi (1970, 1973) and Noguchi and Masuda (1971) have replicated the Jameson and Hurvich study. Despite various changes in experimental conditions to maximize the probability of getting negative slopes, none were observed for any S at any treatment level in Flock and Noguchi's two studies, and only two negative, though slight, slopes were found in Noguchi and Masuda's study. All of these studies consistently showed that slopes systematically decreased as the reflectance of a test region became lower.

Such conflicting results between Jameson and Hurvich's study and its replications strongly suggest that there should be at least two different phenomena or perceptual dimensions, brightness and lightness. The stimulus arrangements and procedures in replicated experiments were engineered so that the Ss could make judgments of brightness without much difficulty. Since Jameson and Hurvich (1961) apparently made no distinction between brightness and lightness, their Ss might have made judgments of lightness rather than brightness. This would be most probable with the black-appearing region having the lowest reflectance. The experiment described below, therefore, was designed to confirm this inference using procedures in which lightness was explicitly set as the phenomenon.

Method

Subjects

Ten Ss, undergraduates at Chiba University, participated in the experiment. They were naive to the purpose of the experiment. All had normal vision in each eye.

Apparatus

As shown in Figure 1, a cross-shaped configuration composed of five Munsell neutral (N) grays was mounted on a Munsell N 9.5 (white) background. It was presented frontally at a distance of 172 cm and was viewed under one of four illumination levels by the right eye through a circular aperture. The three grays in the vertical direction of the cross from top through center to bottom were N 5.0, N 9.0, and N1.5 in Munsell value. The two grays to the left and right of the center square were N 3.0 and N 7.0 in Munsell value, respectively.

![Fig. 1. Six test-field (TF) regions: five squares in a cross and the background of the cross. (Munsell values, contrast ratios, and luminances are shown in Table 1.)](image)

Table 1. Luminances in log cd/m² of TF regions under four levels of illuminance.

<table>
<thead>
<tr>
<th>Location</th>
<th>Log Contrast Ratio</th>
<th>Bottom</th>
<th>Left</th>
<th>Top</th>
<th>Right</th>
<th>Center</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Munsell Value</td>
<td>2.59</td>
<td>1.50</td>
<td>.95</td>
<td>2.34</td>
<td>2.72</td>
<td>2.97</td>
<td>3.04</td>
</tr>
<tr>
<td>Munsell Value</td>
<td>1.61</td>
<td>1.58</td>
<td>1.96</td>
<td>1.36</td>
<td>1.73</td>
<td>1.97</td>
<td>2.08</td>
</tr>
<tr>
<td>.62</td>
<td>-.41</td>
<td>-.66</td>
<td>-.28</td>
<td>.15</td>
<td>.40</td>
<td>.51</td>
<td></td>
</tr>
<tr>
<td>0*</td>
<td>-1.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The illuminance calibrated at this level was approximately 10.5 lux.
Each square of gray, measuring 3 cm on a side, subtended approximately 1 deg of arc and the total angular diameter of the display, including the configuration and visible part of the background, was 7 deg of arc.

A slide projector with an iodide halogen lamp (24-v, 150-w) illuminated the display. The projector was suitably screened to reduce stray light. In order to vary the illumination of the display, four different combinations of neutral density filters could be placed in a holder located behind the eye-piece. The four log relative illuminances were 0, 0.62, 1.61, and 2.59. The luminances of the five square and their background expressed in log cd/m² are shown in Table 1.

**Procedure**

Before being presented the display, each S was trained to establish a subjective scale of lightness, using a Munsell N scale. The E presented nine Munsell patches (3 cm × 3 cm), ranging from N 1.0 to N 9.0 in steps of 1.0 in Munsell value, at random on a black background. The S was asked to rank these nine patches in order to make the Munsell N scale. The S also was told to remember them and to sort into five achromatic color categories: black, dark gray, medium gray, light gray, and white. Then the E presented the five Munsell patches, N 1.0, 3.0, 5.0, 7.0, and 9.0, one at a time, and asked the S to say whether it was black, dark gray, medium gray, light gray, or white, and thereafter to assign it the Munsell value. Again, the E showed the nine Munsell patches one at a time in a random order and asked the S to number them correctly. After this task was completed, the following instructions were given: "In the experiment you are going to participate, a black which is blacker than N 1.0 or a white which is whiter than N 9.0 might appear. In this case you can use a number such as 0 or 10. That is, you can use any number within a 0-10 lightness scale."

After this training period, S was taken to the observation booth where he received instructions about his task and left in the dark for 10 min. The instructions given to the S were:

"Your task is to make judgments about blackness, grayness, and whiteness. A cross composed of five squares will be presented to the right eye. Each square in the cross will be designated by the location, 'top', 'center', 'bottom', 'right', and 'left', and the circular field surrounding the cross will be called 'background'. You are asked to look at a square (or background) designated by its location and to say at once whether it is 'black', 'dark gray', 'medium gray', 'light gray', or 'white', and then to assign a number to it, using the previously learned Munsell scale. For example, say 'black, 1.0', 'medium gray, 5.0', 'light gray, 7.0' and so on. Also, you can use fractions such as 'black, 1.5', 'medium gray, 5.5', or 'white, 9.5'."

In short, in the present experiment, two classes of judgments were made: (1) First the S was asked to look at one of the six test field (TF) regions designated by its location and to say at once its achromatic color (category judgment); and then (2) he was required to assign a number to it, indicating its degree of lightness (numerical judgment).

The E specified the TF region after the display had appeared for 10 sec. The Ss were instructed not to look at any particular area until the E specified the TF region. Each S was tested for each of the six TF regions (5 squares and 1 background) at each of the four illumination levels. The total of 24 treatment-combinations was presented in a random order with a 20-sec dark interval between trials. This procedure was replicated four times with different random sequences. Therefore, each S made a total of 96 judgments each for the two classes of tasks.

**Results and Discussion**

**Achromatic Color Naming**

First look at the results as to how the Ss assigned the five categories of achromatic colors to each of the six TF regions under the four levels of illuminance. Figure 2 shows schematically the distribution of the total of 960 category judgments (6 TF regions × 4 illuminances × 4 replications × 10 Ss). The TF region with the highest contrast ratio (CR=1.50) of surround to focal luminance (or lowest reflectance, N 1.5) was almost always reported to be 'black', regardless of changes in illuminance. Only one out of the total of 160 judgments for
this TF region was named 'dark gray'. Achromatic color namings for the other TF regions with lower CRs (or higher reflectances) were less constant: As illuminance was increased over approximately 2.6 log units, some TF regions shifted from black to gray (mostly 'dark gray') and some from gray (mostly 'light gray') to white. Such shifts of color names were largely restricted within two adjacent region categories: for example, a S who named a TF 'dark gray' under one level of illuminance tended to name the same TF region 'medium gray' under the higher level of illuminance, but he continued to use that same color name with further increases in illuminance. The total of each category distribution over changes in illuminance (with the category responses totaled across the different TF regions) clearly demonstrates that 'black' was virtually invariant and 'white' was very susceptible to changes in illuminance (See the rightest column of Figure 2).

It may be useful to introduce the concept of achromatic color constancy for describing appearances of achromatic colors with changing illumination. It can be defined like this: the achromatic colors of surfaces like the Munsell gray series remain relatively constant over a reasonable range of the photopic vision. That is, a TF region as named 'black' (or 'gray' or 'white') under one level of illumination continues to appear to be 'black' (or 'gray' or 'white') under all other levels of illumination. In fact, Flock (1974) demonstrated that this class of achromatic constancy was almost perfect over changes in illumination of 2.8 log units. His Ss reported that, regardless of changes in illumination, the TF square with the highest CR was black; the TF squares with lower CRs were white; and the TF squares with intermediate CRs were gray. In Flock's experiment there were virtually perfect constancies for all the seven TF squares arrayed in a cross-shape on a white background. In the present experiment, however, the degree of achromatic constancy changed with different TF regions. As stated earlier, perfect constancy occurred only for the TF region of the highest CR that was reported to be black. As the CR of the TF region was made lower, achromatic constancy became less perfect. Such difference may be due to the mode of adaptation: Flock's Ss were light-adapted for 42 seconds to the level near the background luminance on each trial without any dark-adaptation between trials, whereas the Ss in this experiment were light-adapted only for 10 seconds with a 20 second dark-adaptation between trials. In addition to this, Flock's Ss were required to employ only three judgment categories, black, gray, and white, instead of five categories which were used in the present experiment.

**Lightness Functions**

The general configuration of lightness judg-
K. Noguchi and M. Motoki: Achromatic Color Constancy and Lightness Functions

Table 2. Individual and mean linear regression coefficients.

<table>
<thead>
<tr>
<th>Lon Contrast Ratio</th>
<th>1.50</th>
<th>1.12</th>
<th>.73</th>
<th>.34</th>
<th>.10</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>-.18*</td>
<td>.07</td>
<td>.43*</td>
<td>.52</td>
<td>.49</td>
<td>.44*</td>
</tr>
<tr>
<td>#2</td>
<td>-.11</td>
<td>-.23</td>
<td>-.03</td>
<td>.23</td>
<td>.48</td>
<td>.90**</td>
</tr>
<tr>
<td>#3</td>
<td>-.08</td>
<td>-.01</td>
<td>.21</td>
<td>.39*</td>
<td>.55**</td>
<td>.63*</td>
</tr>
<tr>
<td>#4</td>
<td>.03</td>
<td>.04</td>
<td>.11</td>
<td>.52*</td>
<td>.13</td>
<td>.27</td>
</tr>
<tr>
<td>#5</td>
<td>.05</td>
<td>.71**</td>
<td>.46*</td>
<td>.31</td>
<td>.30**</td>
<td>.30*</td>
</tr>
<tr>
<td>#6</td>
<td>.08</td>
<td>.35</td>
<td>.14</td>
<td>.29</td>
<td>.39</td>
<td>.38**</td>
</tr>
<tr>
<td>#7</td>
<td>.09</td>
<td>.00</td>
<td>.30</td>
<td>-.08</td>
<td>.80**</td>
<td>.97*</td>
</tr>
<tr>
<td>#8</td>
<td>.17</td>
<td>.46</td>
<td>.51</td>
<td>.26</td>
<td>.40</td>
<td>.99**</td>
</tr>
<tr>
<td>#9</td>
<td>.32*</td>
<td>.29</td>
<td>.57</td>
<td>.33</td>
<td>.48</td>
<td>.53</td>
</tr>
<tr>
<td>#10</td>
<td>.41</td>
<td>.75*</td>
<td>.16**</td>
<td>.28*</td>
<td>.37**</td>
<td>.41*</td>
</tr>
<tr>
<td>Mean</td>
<td>.07</td>
<td>.24</td>
<td>.29</td>
<td>.31</td>
<td>.45*</td>
<td>.51**</td>
</tr>
<tr>
<td>$r^2$</td>
<td>.82</td>
<td>.75</td>
<td>.81</td>
<td>.80</td>
<td>.96</td>
<td>.98</td>
</tr>
</tbody>
</table>

* $p<.05$        **$p<.01$
appearing TF region; (2) 21 positive functions that were mostly observed for the light gray-
and white-appearing TF regions; and (3) 38 functions with zero slopes or possibly nonlinear
components which will be discussed later.

Mean linear regression coefficients are also
given in Table 2 (one up from the bottom).
ANOVA was run to determine whether these
coefficients changed with different TF regions.
There was a significant main effect, \( F(5,45) = 5.73, p<.01 \). It seemed that the slopes tended
to become increasingly steeper as the CR of
TF region was decreased. However, the slopes
were not significantly different from zero ex-
cept for the two steepest ones obtained for the
TF regions having the CRs of 0 and .10 (N 9.5 and N 9.0).

It should be pointed out that the use of linear
regression coefficients omitted any reference
to possibly nonlinear fluctuations in lightness
judgments over changes in illuminance. The
values of \( r^2 \) in Table 2 suggest the existence
of nonlinear fluctuations at least for the TF
regions of higher CRs. As a rough measure of
the degree of such nonlinear effects, there-
fore, the reversals in each S’s scalar judgments
for the four successively increasing levels of
illuminance were counted. If S’s judgments
for some TF region under the four levels of
illuminance, for example, were 2.0, 2.5, 2.4,
and 2.3, respectively, there would be two
reversals. Had his judgments been 2.9, 3.8, 3.8,
and 2.9, there would be 1 reversal and 1 tie.
Correspondingly, for each TF region with 10
Ss, 30 reversals would be possible. For the 6
TF regions, Bottom (N 1.5), Left (N 3.5), Top
(N 5.0), Right (N 7.0), Center (N 9.0), and Back-
ground (N 9.5), the reversals were 10 (with 3
ties), 9, 9 (with 2 ties), 9 (with 2 ties), 2 (with
1 tie), and 2, respectively. It seemed, therefore,
that the reversals were sufficiently rare to
warrant the use of linear regression coefficients
for Center and Background which were lowest

\(^2\) \( \beta_2 \) was positive only when the background served
as TF. This suggests that it would have been
difficult or impossible for the Ss to separate
between lightness and brightness due to the
appearance of the background which tended to
be seen as a film color rather than as a surface
color. So, it may be that the Ss might have
responded to brightness instead of lightness.

in CR. For the other TF regions of higher
CRs, however, there did exist nonlinear effects.

Closer inspection of the data configuration in
Figure 3 also reveals that there would be
nonlinear fluctuations of lightness judgments
over the change in illuminance. Namely, these
seemed to exist two components to the lightness
functions: in the case of focal TF regions
(squares), there was an initial, pronounced in-
crease in lightness for the lower levels of illu-
minance and then slight decrease for the highest
level of illuminance; and when the background
served as a TF region, this relation was re-
versed. Now it is clear that the fit of the linear
relationship will not be satisfactory. There-
fore, the fluctuations of lightness judgments
for each TF region over the four illuminances
were fitted by least-squares curves (second-
degree). The quadratic equations and their
coefficients of determination \( \left( R^2 \right) \) are listed in
Table 3.

Table 3. Curvilinear regression (second-degree)
equations \( y = \beta_0 + \beta_1x + \beta_2x^2 \) and coefficients of
determination \( \left( R^2 \right) \).

<table>
<thead>
<tr>
<th>TF Region</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom</td>
<td>.841</td>
<td>.183</td>
<td>-.043</td>
<td>.974</td>
</tr>
<tr>
<td>Left</td>
<td>1.981</td>
<td>.754</td>
<td>-.196</td>
<td>.998</td>
</tr>
<tr>
<td>Top</td>
<td>3.451</td>
<td>.701</td>
<td>-.158</td>
<td>.937</td>
</tr>
<tr>
<td>Right</td>
<td>5.102</td>
<td>.853</td>
<td>-.210</td>
<td>.998</td>
</tr>
<tr>
<td>Center</td>
<td>7.103</td>
<td>.747</td>
<td>-.113</td>
<td>.998</td>
</tr>
<tr>
<td>Background</td>
<td>7.275</td>
<td>.284</td>
<td>.108</td>
<td>.998</td>
</tr>
</tbody>
</table>

The values of \( R^2 \) prove that the match
between the observed data-points and the values
predicted from these equations is rather close
and, therefore, that lightness functions are
curvilinear for a semilog transformation.
Furthermore, it can be seen from the sign of the
\( \beta_2 \) coefficients that the lightness functions
for the TF regions which are five squares
forming a cross would be regarded as positively
decelerating curves (concave downward) and the
function for the TF region which is a
background surrounding the cross is negatively
accelerating curve (concave upward).

**General Discussion and Conclusions**

Several theories of brightness perception
under contrast (Bartelson & Breneman, 1967;
Jameson & Hurvich, 1964, 1970; Marimont, 1962; Stevens & Stevens, 1960) predict that brightness judgments decrease rather than increase for certain conditions when illuminance is increased. These predicted negative functions seem to occur when the CR of surround to focal luminance is relatively large and is held constant over the increases in illuminance. As the magnitude of the CR is decreased, functions correspondingly become less negative, become flat, and then become increasingly positive. It is not clear, however, whether these theories really refer to the phenomenon of brightness. It seems probable that the data supporting some theories have nothing to do with brightness and instead would have been obtained under conditions in which only lightness judgments could be made.

The opponent-process theory (Flock, 1970; Jameson & Hurvich, 1964, 1970) can be interpreted to tie together brightness and lightness judgments: As illuminance is increased, all light grays and whites will become increasingly "brighter" and "whiter" at the same rate, all dark grays and blacks will become increasingly "less bright" and "blacker" at the same rate, and some mid-gray will not change either in brightness or in lightness. This type of theory unfortunately blurs any distinction between these two perceptual dimensions. It is not clear from Jameson and Hurvich's (1961) use of the terms brightness and lightness if they made a distinction. Later studies (Flock & Noguchi, 1970, 1973; Noguchi & Masuda, 1971) where the S's task was explicitly to judge brightness have all failed to find significant negative or zero functions.

In the present experiment, therefore, it was questioned whether negative or zero functions could be observed when the Ss were required to make lightness judgments instead of brightness judgments. In order to facilitate lightness responses, the task of achromatic color naming was given to the S before making lightness judgments in terms of the Munsell neutral scale. Although most of the lightness functions were positive or approximately zero, negative functions, though very exceptional, were also observed for some Ss when the TF regions had the largest CR. In his lightness match experiment, Helson (1943) also found approximate zero (.05) and negative (−.12) functions for the TF regions of the largest CR. These findings suggest that the possibility of finding negative and zero slopes seems to be higher for lightness than for brightness judgments. It should be pointed out, however, that the significant negative slope was contributed by only one of the 10 Ss in the present experiment and was much less steep than that of Jameson and Hurvich's (1961). Furthermore, mean lightness judgments averaged over all the Ss showed no negative slopes even for the TF region of the highest CR which gave an almost zero slope. The previous studies (Flock, Noguchi, & Muter, 1974; Kozaki & Noguchi, 1976) on lightness judgments also demonstrated that the occurrence of negative slopes was very rare. The negative slopes obtained in Helson (1943) and Jameson and Hurvich (1961) would probably be due to this: in both experiments, there was an anomalous asymmetry between preadaptation and TF luminances; their Ss were allowed to look back and forth from the brightly illuminated comparison region to the very dimly illuminated TF region. In other words, the S's eyes were always adapted to a relatively high level of illumination when the S observed the darkest TF region under a very low level of illumination. It would seem reasonable that the negative slopes obtained under such anomalous conditions are rather exceptional. It can be concluded, therefore, that the theoretical claim for the generality of negative functions for both lightness and brightness domains should be rejected.

The results from lightness and brightness experiments (for lightness, Flock et al., 1974; Kozaki & Noguchi, 1976; this study, and for brightness, Flock & Noguchi, 1970, 1974; Noguchi & Masuda, 1971) seem to indicate a similar data configuration when plotted against log relative illuminance: there would be a high positive correlation between a group of lightness functions and those of brightness functions. As pointed out by Flock (1974), however, the presence of a correlation between lightness and brightness should not be allowed to blur the difference between these two dimensions. It is suggested that given a log transformation of illuminance, lightness functions are concave downward, and brightness functions are concave.
upward: as illuminance is increased, there is a small amount of change in the lightness of a TF region and the decelerating use of number in judging its grayness, whereas there is an accelerating use of number in specifying the brightness of that TF region.

References


— Received March 20, 1982 —